Requested Patent: WO0159612A2

Title: IMPROVEMENTS RELATING TO DATA FILTERING ;

Abstracted Patent: WO0159612 ;

Publication Date: 2001-08-16 ;

Inventor(s): GIBBS JASON (GB) ;

Applicant(s): ADSCIENCE LTD (GB); GIBBS JASON (GB) ;

Application Number: WO2001GB00603 20010214 ;

Priority Number(s): GB20000003382 20000214 ;

IPC Classification: G06F17/30 ;

Equivalents: AU3210501, GB2369210 ;

ABSTRACT:

A client side data processing apparatus configurable for use with a computer system having a browser and a method of operation thereof is configured to process a block of information requested and received over a telecommunications network. The information comprises potentially required data content and tag type mark up language commands for controlling the display of the potentially required data content by a given browser. The apparatus comprises means for identifying a plurality of types of received tag type commands; and means configurable to process the received and identified tag type commands according to a pre-defined set of rules. Processing may include identification of particular data types within the read content data and processing of any identified content data accordingly, such as for example via matching with a key word list.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number
WO 01/59612 A2

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number: PCT/GB01/00603

(22) International Filing Date: 14 February 2001 (14.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0003382.9 14 February 2000 (14.02.2000) GB

(71) Applicant *(for all designated States except US)*: **AD-SCIENCE LIMITED** [GB/GB]; Cray House, Woodlands Road, Henley-on-Thames, Oxfordshire RG9 4AE (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **GIBBS, Jason** [GB/GB]; 31 Market Place, Henley-on-Thames, Oxfordshire RG9 2AA (GB).

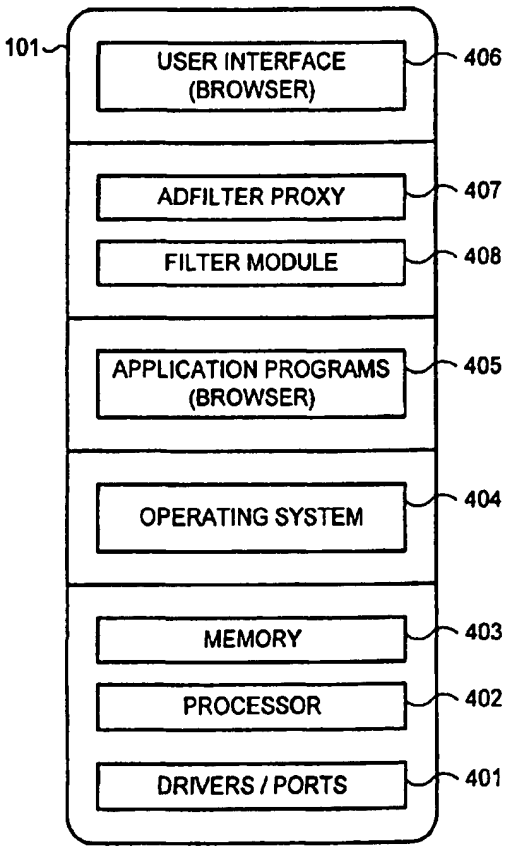
(74) Agents: **AHMAD, Sheikh, Shakeel** et al.; David Keltie Associates. 12 New Fetter Lane, London EC4A 1AG (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: IMPROVEMENTS RELATING TO DATA FILTERING



(57) Abstract: A client side data processing apparatus configurable for use with a computer system having a browser and a method of operation thereof is configured to process a block of information requested and received over a telecommunications network. The information comprises potentially required data content and tag type mark up language commands for controlling the display of the potentially required data content by a given browser. The apparatus comprises means for identifying a plurality of types of received tag type commands; and means configurable to process the received and identified tag type commands according to a pre-defined set of rules. Processing may include identification of particular data types within the read content data and processing of any identified content data accordingly, such as for example via matching with a key word list.

WO 01/59612 A2

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

IMPROVEMENTS RELATING TO DATA FILTERING

Field of the Invention

The present invention relates to improvements in the field of data filtering and particularly although not exclusively the invention relates to filtering information obtained over a telecommunications network such as the Internet and World Wide Web. The invention also relates to filtering advertising information and information of an adult nature such as pornography, bad language, violence/suicide and drugs.

Background to the Invention

With the World Wide Web (WWW) growing and projections for web users increasing exponentially, the concerns among individuals and corporations as to use and abuse of information available on the web is growing rapidly. Increasingly adult material is almost impossible to escape, adverts are becoming more focused and intrusive and privacy is being abused.

To view information on the World Wide Web or Internet it is known to equip an Internet terminal, such as a personal computer, with means for accessing the Internet and World Wide Web, this means being known as a browser. The vast majority of information obtainable from the World Wide Web or Internet is written in hypertext mark – up language (HTML) which is a strictly defined method of presenting textual material intended for use in the World Wide Web. HTML enables control of page layout and format of characters and provides for inclusion of active links. Such active links contain a universal resource locator (URL), a URL being an address used to specify the location of a multi-media document in the World Wide Web.

By specifying a URL any HTML page stored electronically on the web can be obtained by a given user and by virtue of the links various other HTML pages can be embedded therein and appear to a given user when not necessarily required. Advertisements, in particular, may appear to users of the World Wide Web in a manner which was not specifically requested. Such material, for

example, comes in the form of banners, pop-up advertisement windows which appear whilst the user is browsing the Internet and is known as "Spam". Pop-up window advertisements require the user to close the relevant windows before continuing. The action of closing a pop-up window advertisement can frequently cause the launch of yet another pop up window and yet more pop-up windows which can waste a given user's time and pop-up undesirable material such as material of a pornographic nature for example.

Access via the Internet to special interest sites and company websites is typically via an Internet Service Provide (ISP). In the past, users had to pay a subscription to an ISP for access to the Internet. However, in recent times this model in certain circumstances has been overtaken by dropping such charges in favour of alternative revenue sources, leading to low or no cost Internet access within the context of a broader Internet commerce-based consumer economy. ISP's have been forced to look elsewhere for revenue, with the obvious alternative source being advertising.

Advertising is now prevalent amongst ISP's and a major source of funding for websites. The increasing sophistication of advertisements results in the majority of web pages containing more marketing material than actual information required by many people. Advertising graphics are generally very extensive in relation to the memory space they take up, slowing download times and substantially increasing the length of time people must stay on line. This results in higher ISP charges (when applicable), higher telephone charges to the user (when applicable) and a waste of user time. Some of the more popular and more frequently visited websites carry so much advertising, relative to the actual content being sought by the visitor, that almost 90% of the download time taken to see the page is the result of advertising content and not the information required. This can be extremely annoying to the given user browsing the Internet and web.

Various additional problems have arisen from the availability of information over the internet and World Wide Web including, for example, a larger amount of pornographic material now available which may be desirable to prevent access to across a whole culture or for an individual family or member of a family etc. In addition, it is now common that employers are keen to restrict access to adult sites and some theme sites such as violence and bad language because of disruption in the workplace and the cost to the employer in terms of wasted employee time and so on.

Yet a further problem with current Internet browsing includes lack of privacy as regards to the sites actually visited by a given user of a browser. Thus, various marketing companies are able to track which websites a given user visits and therefore compile statistical information or use the information detrimentally to the user. There is thus a need to improve privacy of a given user's actions on the Internet and World Wide Web so that sites visited can be prevented from becoming known to trackers and market researchers.

As indicated above yet a further problem associated with current Internet usage is the time taken to download relevant required information that a given user has requested, the time taken being considerably extended by virtue of the desired content being entangled with advertising material which lengthens download times considerably.

Although it is desirable to filter advertising information, pornographic data and the like and also to improve privacy of a given user's choice in sites visited it is also a problem that existing prior art web filters, as far as the inventors are aware, do not allow the filtering to be turned off for whole sites or for individual pages within a site if the user so requires. In other words it may be desirable for a given user to allow a certain amount or type of advertising to be allowed through when browsing.

Prior art web filters to date fall into two categories: Those in the general web filter software market include stand alone programs to prevent advertising material and the like getting through and include the following systems, listed by their trade names as follows:

- Web Washer™, Intermute™, Internet JunkBuster and AdWiper™

In the adult filter market various systems exist, again listed by their trade names, as follows: Cyber Patrol™, Net Nanny™, Surf Watcher™ and Surf Control™.

None of the above prior art web filter systems are integrated into a single product which caters for filtering both subject matter of an adult nature and advertising material. Additionally, there is a lack of facility with regards existing web filter systems in terms of providing users with a means for reporting advertisement types missed by existing web filters.

As indicated above a variety of prior art web filters are known. Thus, for example international patent publication no. WO 97/49252 (Manickavasagam) discloses a medium manipulator which may be used to manipulate various media objects requested by a given client's request and in particular discloses a method of calling service devices to perform data compression or pornography detection on particular images. Detection of images such as pornographic images is described by way of statistical analysis of colours in a given image such that should a given percentage of flesh tone colours appear in the picture the image may be prevented from display as being likely to be of a pornographic nature. Such a method is configured to analyse image data and not text based material and therefore is susceptible to missing pornographic material of a textual nature.

An alternative prior art web filtering system and method is disclosed in US Patent no. US 5987606 (Derosa) which works on a known principal of searching a list of allowed or excluded web site addresses. The stored list stores a list of

URL's which the system searches for, a given URL being identified in an incoming HTML page, and should a match be made then the incoming page, or at least a part of the data contained therein, is manipulated so that it is either made non-visible or removed completely. A problem with such a system is that there is a requirement for a team of Internet/World Wide Web searching staff to identify relevant web page addresses which are to be effectively rejected. In view of the vast amount of new advertising material and pornographic material appearing on the World Wide Web and the Internet such teams are only partially effective since vast numbers of such addresses will be missed by virtue of the sheer number making it an almost impossible task to keep such lists up to date. Additionally, a given web user has no, or little control over the exact selection of web site addresses actually excluded or manipulated in some appropriate way.

In view of the above there is clearly a need to improve web filters such that advertising material, adult type material and the like can be identified in a more reliable manner and be relieved from reliance upon image analysis and approaches utilizing out of date lists of relevant URL addresses identified by specific teams of people policing the World Wide Web and Internet. Accordingly, it is an object of the present invention to address at least some of the above described problems.

Summary of the Invention

According to a first aspect of the present invention there is provided a client side data processing apparatus configurable for use with a computer system having a browser, the apparatus configured to process a block of information requested and received over a telecommunications network, the information comprising potentially required data content and tag type mark up language commands for controlling display of the potentially required data content by the browser, the apparatus comprising: means for identifying a plurality of types of the received tag type commands; and means configurable to process the received and identified tag type commands according to a pre-defined set of rules.

The consideration of tag type commands provides an extra dimension to the filtering processes which has not been available before. The problems of using up-dated URL lists of banned web sites are mitigated because the content of the page is being considered rather than a possibly out-of-date data descriptor. The present invention is particularly useful for recognising advertisements. Typically, advertisements will have to use these tag type commands to position themselves appropriately within the page to be displayed or will have special advertisement type features such as blinking or referral to another web site. The specific types of commands can be detected and appropriate processing can be carried out, usually in the form of filtering but it is also possible for tag type commands to be replaced with more appropriate ones or for them to be modified in some way to make them more suitable. These further processing instructions are determined by the above mentioned rules.

The identification means may comprise means for identifying a plurality of types of the tag type commands which are used for controlling display of electronic advertising information. Knowledge of all of the types of commands used for advertising provides a difficult to bypass screen which can be used to delete all such recognised advertisement data if required.

Preferably the identification means comprises means for reading the received information character by character and means for comparing a pattern of the characters with a pre-stored list of tag type command syntax. This enables recognition of tag type commands to be achieved in a simple way. Recognition of a tag type also determines what further processing is to be carried out.

More specifically the identification means may comprise means for identifying tag type commands which specify a specific size of an electronic data banner to be displayed. As most advertising uses standard banner sizes, this provides an fast and effective way of identifying potentially undesirable content in the received information to be displayed.

The set of rules may be adaptable according to a given user's requirements of the apparatus. This tailoring of how the filter is to function enables it to be updated with new information regarding developments in the tag type commands and also to be flexible to changes in user requirements of the apparatus.

The rules may specify the processing to include: modifying an identified tag type command according to a pre-defined criteria thereby changing its effect on execution; removing the tag based command from the received information; allowing the identified tag type command to be executed without any modification; or replacing the identified tag type command with a stored tag type command. These different options again provide a given user with the ability to vary the effects of filtering on the received information in different ways such that advantageously user-desired results can be achieved. Also the configurable processing means may be modified according to a given user's preferences to support further the user adaptability of the apparatus.

The flexibility provided by the present invention allows the filtering to be turned off for whole sites or for individual pages within a site if the user so requires, such that a certain amount or type of advertising can be allowed through when browsing. This is readily achieved, for example, by the rules specifying differences in processing in dependence on the URL of the web site being visited.

Preferably the means configurable to process the identified tag type commands according to a pre-defined set of rules includes further processing means configurable to search for potentially non-required pre-defined data types in the data content associated with the identified tag type. This provides a higher degree of resolution in the capabilities of the apparatus because in addition to using tag type commands, the content associated with those commands can also be checked for non-required data identifiers. Also this enables every meaningful part of the received information, namely tag types commands and content

relating to those commands to be searched and used in further processing of the received information, typically selective filtering.

Another fundamental advantage of considering both the command tags and data content of received information is that whole web sites can be filtered on the basis of their content regardless of what they are called, namely their URL. In this way, new sites which arise can be filtered even if their URLs have not previously been known.

In exemplary embodiments of the present invention, the way in which command tag and content filtering is achieved is for the further processing means to comprise: means for reading the data comprising the content; means for comparing the read content data with a stored list of potentially non-required data types to search for in the content data and identifying any matches found; and means for processing the identified matched data in accordance with previously stored processing instructions associated with each the potentially non-required data type in the list.

The list of stored potentially non-required data types may comprise a list of human language words or a list of certain groupings of human language words. These are words which are looked for in the content and which if found are indicators that the content relating to a particular tag type command is of a given nature which it may be desired to filter. In most cases, the means for processing the identified matched data includes means to prevent the display of the identified matched data.

According to a second aspect of the present invention there is provided, in a computer system having a browser for displaying requested information received over a telecommunications network, the information comprising potentially required data content and tag type mark up language commands for controlling the display function, a method of controlling the functionality of the received tag type commands, the method comprising: identifying a received tag type

command; and processing the identified tag type command according to a pre-defined set of rules configured for application to tag type commands of the identified type.

Preferably the step of identification comprises comparing the received tag type command with a pre-stored list of tag type commands and identifying a match. If the match cannot be found, details of the tagged command under consideration may be saved and a warning message may be provided to the user of the system. This enables future proofing of the apparatus as the receipt of a new type of command will be flagged to the user and appropriate action to incorporate its details can be taken.

Suitably, following a given tagged command having been identified, the step of processing includes loading executable processing instructions associated with the tag type and executing the processing using the tag type accordingly. These instructions reflect the user's pre-determined way of dealing with each particular command. The user may configure the apparatus to carry out very different instructions in dependence upon the type of tag type command that is identified and this provides more user flexibility in the apparatus.

For example, the processing step may include selecting one of: ignoring the tag type command; enabling the tag type command to execute in its original form; but replacing the tag type command with a pre-set replacement command; and modifying the tag type command according to pre-defined stored rules for the tag type command thereby changing the executable effect of the identified tag type command.

The processing step in an embodiment of the present invention includes: reading the data comprising the data content associated with the tag; comparing the read content data with a list of previously stored potentially non-required data types to search for in the content data and identifying any matches found; configuring processing means with previously stored processing instructions

associated with each potentially non-required data type in the list; and processing one or more the identified matched data in accordance with the associated instructions.

According to a third aspect of the present invention there is provided a method of filtering non-required data from information received over a telecommunications channel, the method comprising identifying a received tag type command from within the received information, scanning the data content specifically associated with the received tag type command and filtering at least a portion of the data content in response to matching an item in a pre-stored data list with the portion of the data content.

By looking at the specific content of a web page rather than just its URL or even just its HTML tags, for example, it is possible to provide a more intelligent filter. The combination of using tag type commands together with the content associated with the commands enables, for example, comments regarding the web page or textual wording to be displayed on the web page to be analysed for filtering purposes. The use of look-up tables for the pre-stored lists advantageously enables fast checking. Finally, whole web sites can be filtered by the simple identification of the beginning and end command tags of a web page and the matching of subject matter within the page with pre-stored non-allowable subject matter; a match indicating all of the content between the start and end command tags needing to be filtered, i.e. the whole web page.

There are preferably a plurality of pre-stored data lists each being specifically associated with at least one tag type command and the matching step preferably comprises searching those lists associated with the received tag type command. Again, by specifying which lists are associated with which tag type commands, only a subset of the possible lists are searched and this means that the checking can be carried out far more rapidly than if all of the lists had to be checked each time.

The filtering step may comprise selectively filtering the portion of the data content without filtering the entire data content associated with the received tag type command. This advantageously enables a very high resolution and intelligent filtering to be achieved because within a received page of HTML, only those aspects need be filtered that cause difficulties, without the need to filter the whole page. An example of where this may be useful is in medical circles where there may be references to parts of the human anatomy which may under some circumstances otherwise lead to the whole page being filtered.

Brief Description of the Drawings

For a better understanding of the invention and to show how the same may be carried into effect, there will now be described by way of example only, specific embodiments, methods and processes according to the present invention with reference to the accompanying drawings in which:

Fig. 1 schematically illustrates the environment in which the present invention may be used as configured to be operated on a client computer system 101;

Fig. 2 shows a typical HTML page of information of the type received over a telecommunications network such as the Internet following a request for the HTML page by a client computer or other terminal as configured in accordance with the present invention;

Fig. 3 schematically illustrates the formatted HTML page illustrated in Fig. 2, the page having been received and processed by a browser operated by the computer processor of client 101 in Fig. 1;

Fig. 4 schematically illustrates components of computer 101 shown in Fig. 1, the components including various standard components such as an operating

system and also filtering components as configured in accordance with the present invention, the filtering components including a data receiving module 407 and a data filtering module 408;

Fig. 5 schematically illustrates, in accordance with the present invention, n filtering (targeter) units as configured on computer 101;

Fig. 6 further details typical attributes associated with a targeter of the type identified in Fig. 5;

Fig. 7 schematically illustrates a key word list of the type associated with the targeter detailed in Fig. 6.

Fig. 8 schematically illustrates a second list of key words; in the form of associated words, of the type referred to by the targeter detailed in Fig. 6;

Fig. 9 further details the main steps executed by the data receiving module 408 of Fig. 4 for passing data received to the filtering module 409 in Fig. 4;

Fig. 10 further details an exemplary sequence of processing steps involved in filtering data received over the Internet as processed by filter module 408 following receipt by data receiving module 407 and comprises a step 1007 of processing located portions of data requiring processing; and

Fig. 11 further details a preferred exemplary sequence of steps involved in the processing step 1007 of Fig. 10.

Detailed Description of the Best Mode for Carrying Out the Invention

There will now be described by way of example the best mode contemplated by the inventors for carrying out the invention. In the following description numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent however, to one skilled

in the art, that the present invention may be practiced without limitation to these specific details. In other instances, well known methods and structures have not been described in detail so as not to unnecessarily obscure the present invention.

In this specification, the terms "filter" and "filtering" refer to processing of data received over a telecommunications network. The filter processing may include modifying some or all of the received data in some way, deleting some or all of the received data, replacing some or all of the received data with other data or in the case of a received command with another command, and allowing the received data to be processed and displayed by a browser in its original (unchanged) form after having been checked.

By data it is meant that information received over a telecommunications network following, for example, request by a client computer or other suitable terminal connected to a network such as the Internet. A response to such an information request may typically include potentially required data and potentially non-required data.

Fig. 1 schematically illustrates a typical environment in which the present invention may be utilized. Thus, a personal computer or networked computer 101 may be configured with electronic processing circuitry in accordance with the present invention or alternatively and in the best mode contemplated the relevant processing may be configured in software. Computer 101 may suitably comprise a processor and memory and all the usual ports and features commonly associated with such computer systems. Thus, computer 101 is provided with monitor 102 having screen 103 and is also provided with input devices such as keyboard 104 and mouse 105. Computer system 101 may be operated by one or more users of the system who wish to obtain information from the Internet and World Wide Web 106. Computer 101 may access Internet 106 via Internet Service Provider server 107 and is configurable to request information from a plurality of distant servers 108 and 109. Computer 101 connects with ISP 107 via telecommunications link 110 through which request messages, known as fetch or

get messages, and receive messages are transmitted electronically. Computer 101 may be invoked to send a required information request to the Internet 106 via a user operating a suitably configured browser as viewed on screen 103 and executed by the computer processor of Computer 101. Suitable browsers include Microsoft Internet Explorer ™ or Netscape Navigator ™ for example. Typically an information request will be generated, under control of a user; by the browser and thereafter transmitted to Internet Service Provider 107 via communications link 110 and whereafter the particular server receiving the request, such as for example server 108, will respond accordingly and transmit the requested information back to computer system 101. Commonly the information transmitted following a request is transmitted using a mark-up language such as hypertext mark-up language (HTML). Upon receiving requested information typically the client side browser appearing on screen 103 is configured to process the incoming HTML page and display it in accordance with formatting commands formed as part of the make up of the HTML page. Pages that point to other pages are said to use "hypertext", this being frequently used in the electronic advertising industry. Thus, commonly for a given HTML page of a website having a substantial audience, the owner of the site can effectively sell links to advertisers such that when a given user requests the page the user also inadvertently receives linked pages comprising advertising material and/or various other kinds of material. Frequently such advertising and additional material is not required by the user making the particular request and the presence of this potentially non-required information may considerably slow down the speed of obtaining any actual required information.

Web pages are most commonly written in a mark up language such as HTML. HTML allows web pages to be produced that include text, graphics, and pointers to other web pages. Each given web page is assigned a URL that effectively serves as the page's worldwide name. By mark up language as used above it is meant a language for describing how documents are to be formatted following their transmission over a telecommunications network in response to a

user's request. Mark up languages, such as HTML for example, thus contain explicit commands for formatting.

A typical HTML page syntax 201 is illustrated in Fig. 2. The HTML language contains explicit commands for formatting as do a variety of other mark up languages. The basic layout of the HTML document 201 is such that a proper web page consists of a head and body enclosed by the strings <HTML> and </HTML>, known as tags, 202 and 203 respectively. Tags are effectively formatting commands, usually in pairs, and the next set in the figure comprises the head tag 204 and its corresponding end tag 205. Tags 202 and 203 declare the web page to be written in HTML and tags 204 and 205 (head) contain a description of the HTML page. The information comprised within tags 204 and 205 is known as meta information and is not actually displayed. In the example shown head tag pair 204, 205 surround meta information 206, the meta information comprising title tag pair 207 and 208 which control display of information 209. In the example shown information 209 simply comprises a given company's name "No1WebFilter, Inc". Following the title portion of the HTML page there is a further component of the page known as the body which is surrounded by BODY tags 210 and 211 respectively – these tags de-limit the page's body which is generally indicated in the figure by vertical parenthesis 212. Within the body the next line comprises a first line surrounded by heading tags (<H1>) 213 and 214 (</H1) respectively – such tagging effectively displays the contents within tags 213 and 214 as the title of the HTML page to be displayed. A similar heading is shown lower down the page by header tag pair (H2) 215 and 216 respectively. The HTML line of code at 217 uses the tag "< IMG SRC = ..." >" which designates loading of an image – in the present case an image from the World Wide Web site www.Webfilter.com – this coding is configured to retrieve an image and thus a user having requested page 201 will also receive the image specified at HTML line 217. Such an image may or may not be required by the user and could unduly increase the time required to download HTML page 201. Various other more simple tags are used in the example such as tags for indicating bold print (B) 218 and 219 respectively. A further interesting

feature shown in Fig. 2 is the use of a so called hyperlink command line 220 which uses the tag pair "< A HREF and </A >" as indicated at 221 and 222 - such a tag pair known as an "anchor" thus defines a hyperlink. A hyperlink is a string of text that is a link to another web page and typically these may be configured to be highlighted on a displayed HTML page in some way such as by using underlining tags 223 (<UL>) and 224 (</UL) surrounding hyperlink 220. Hyperlink 220 comprises a link to the home page of the company using the URL "webfilter.com" which may enable a user to click on the resulted formatted printed information "New Filter" to be printed on the user's browser at the particular position on the screen of a given monitor or other display device being used.

The HTML page detailed in Fig. 2 is provided for illustrative purposes and the resultant formatted page actually observed on a given user's screen following a request for the page is schematically illustrated in Fig. 3. Various features discussed above can be observed on screen 103 such as the formatted page 301 and for example the underlined hyperlink as displayed at 302 and as discussed above. Additionally, the image called at line 217 in Fig. 2 is displayed at 303 and comprises the company logo.

There are a large selection of common HTML tags which can be reviewed in a wide variety of references such as for example those published by The WillCam Group and Gregory Consulting. Most tags are paired, but some are singular in the HTML standard. An example of the use of a singular tag is the start of paragraph tag <P> as for example used in Fig. 2 at 225.

From the above description it is therefore clear that a typical HTML page comprises many pairs of tags and singular tags, but in all cases the body of the page comprises HTML mark up tags to effect the format of required text and images to be displayed, both the required information text/images and the mark up (formatting) text being present within the body.

Many prior art filtering mechanisms utilize referral to lists of URL's which a team of workers has compiled as being unsuitable for reasons of content of pornographic nature and the like. This approach as discussed in the introduction requires the list to be maintained at considerable expense and furthermore provides highly limited protection from potentially unwanted downloaded material inadvertently requested through embedded HTML hyperlinks etc, because typically such prior art filters will only search the meta information so as to determine whether or not a page contains potentially non-required material.

Some of the material specified in the HTML page will be required by a given user requesting the page. Thus for example, although only for illustrative purposes, the HTML page illustrated in Fig. 2 may be required by a given user who thereby receives the information as shown in Fig. 3. However, if the page was configured to comprise advertising material in place of logo 303 for example then the resulting advertisement image may in fact not be required by the user and therefore be filtered in some way if processed in accordance with the apparatus and methods of the present invention. Similarly certain text, such as for example, that shown at 304 could be filtered if configured to be processed in accordance with the apparatus of the present invention.

In contrast to known prior art web filters the present invention provides a lower level of operation for determination of whether or not a requested page comprises potentially non-required information content. The present invention utilizes a plurality of filters (known as targeters) which may be configured as programmable objects used to search for patterns of one kind or another within a block of HTML text. These targeters are described in greater detail later. Typically new forms of tag pairs may be used by advertisers so creating new techniques to place their adverts (ads). The present invention enables simple creation of new filters to detect such tags and thereby remove any such new advertisements. If configured in software the filtering engine of the present invention may thus be simply modified by incorporation of a new line or two of text to the relevant filtering engine file which may thereafter be compiled and

executed to filter such unwanted advertisements in a pre-configured manner according to specific settings set by a user and/or a manufacturer of the advice.

Referring now to Fig. 4 the components of client computer 101, which includes a filtering engine, as configured in accordance with the present invention are schematically illustrated. Computer 101 as configured in accordance with the present invention comprises various standard components such as drivers/ports 401, processor 402, memory 403, operating system 404, application programs 405 and user interface (browser) 406. Browser 406 may be invoked in the usual manner and executed for use. Computer system 101 additionally comprises components of the present invention which include filtering components 407 and 408 respectively. Data receiving module 407, called in the present example, the add filter proxy 407 is configured to receive requested information and initialize the filter module 408 so as to enable module 408 to filter the specific block of HTML under current consideration for processing. Thus, in operation a block of HTML data is received by the data receiving module 407 and thereafter passed to filter module 408 for processing in accordance with methods of the present invention.

The invention utilizes a plurality of filters or targeters as illustrated schematically in table form in Fig. 5. Each targeter is configured in software and is called as a sub-routine of filter module 408. In the illustrative example shown a plurality of targeters are identified at column 501 and their names/corresponding parameters detailed in column 502. Effectively upon a call being made to a given targeter its stored parameters may be incorporated into filtering engine 408 to enable required processing to be undertaken. A series of targeters are shown such as for example targeter no. 1 at 503 known as an anchor targeter 504. Further targeters may be configured in engine 408 such as targeter no. 4 at 505 and targeter number n at 506. Each targeter may be considered to represent a sub-filter for particular processing required upon detection of a given type of mark-up language tag. Thus, for example targeter no. 1 at 503 corresponds to processing required when filter module 408 detects the presence of an anchor

tag defining a hyper-link as discussed in relation to Fig. 2 above. Similarly, targeter no. 4 in the present example may be invoked upon detecting tags associated with a pop-up type window of a first type A. Many such targeters can be configured each corresponding to a particular tag structure defined in HTML.

Fig. 6 schematically illustrates a set of parameters that are associated with a given general type targeter M as indicated at 601. The parameters of targeter M are comprised or stored electronically within memory at 602 and comprise a series of pre-defined operating parameters specific to the particular targeter. Thus, for example at rows 603, 604, 605, 606, 607 and 608 in the illustration, stored parameters are respectively stored as pre-defined parameter values 609, 610, 611, 612, 613 and 614. In the present example parameter 1, defines the relevant begin and end tags for the given targeter, parameter 2 defines certain disallowed characters, parameter 3 defines certain key words of a first type and parameter 4 describes certain key words of a second type. These key words are stored in filter lists and their parameters 2 and 3 are simply calls to these lists (see Figs 7 and 8 later). By key words it is meant pre-defined words or character strings which the targeter is to be configured, during operation, to detect from within information received in the form of an incoming HTML page. The targeter 601 also has a reference at parameter n-1 607 to relevant registry sizes tables 613 (otherwise not shown). These provide format size information (non-textual) which may be required to be considered to determine whether or not to filter this tag. Also, certain rules concerning modification of targeted text in a predefined user specific way may be provided such as the rule (pre-programmed instruction) stored in row n, 608. The particular configuration of a given targeter will depend upon the nature of the type of data held within the particular command tags of concern.

As a specific example the anchor targeter 504 (targeter no. 1 in Fig. 6) is configured with the following parameters so as to effect processing of a detected hyperlink.

Referring to Fig. 4, upon data receiving module 407 receiving an incoming HTML page it passes, to filter module 408, information comprising the name of the host, the current URL and the name of the referrer. The full URL is broken up at this initial stage as its components are required as parameters for the subsequent targeter calls. The filter module 408 thereafter initializes each of its targeters including the anchor targeter 504 with this information such that each targeter need not request this information individually. Thus, in the case of the anchor targeter 504 being executed by filter module 408 it is required to be prior configured with certain rule-based parameters of the type indicated in general in Fig. 6. In particular, in the case of the anchor targeter, module 408 is thus pre-configured with a set of parameters and therefore effectively knows, for example, the following:

Begin-tag = "<A"
End-tag = "</A>"-
Not to allow other characters within the Begin-tag or End-tag text.
The character preceding the Begin-tag must not be a quote character.
The character preceding the Begin-tag must not be an A-Z character.
The current Host must not be contained within the Begin-tag/End-tag block.
The End-tag is all that is needed to satisfy the end of the target.
To use a keyword list stored in a section of the Filter's memory.
To search for an inner target bound by the tags "<IMG" and ">".
To also detect image Height and Width values stored in the particular sub-file of the Filter's file concerning particular advertisement sites to filter out.
To signal found when a match is found by Keyword or Size.
To apply keyword detection to the outer target ("<A...</A>").
To apply modification to the inner target ("<IMG"...">").
How to modify its block of target text.

As illustrated above the invention also utilizes key word lists - that is simple lists of indexed words encrypted and stored in the Windows™ Registry or another

suitable memory area. In the example given above, nesting of target detection can be configured such as "To search for an inner target bound by the tags "< IMG" and ">". Additionally key word lists can be used to modify the operation of a given targeter. Such key word lists are stored in a specifically configured file of filter module 408 and indexed by one or more specifically configured targeters. Such key word lists may suitably be configured to call sub-filters stored in the same format. Similarly, certain key words may be utilized to invoke activation of a given targeter upon their detection on whether or not the given targeter is currently being executed. For example, if a given website name is read by a currently executed targeter of filter module 408 then this may be present in a given key word list to which the current targeter relates and therefore be utilised to invoke a further targeter and process the detected data content accordingly. Thus, for example, suppose that the website "Mid Farm" displays anchor tag based advertisements from the company Flycast, then the entry into the key word list for the anchor targeter that activates the required anchor targeter may be configured as:

"89" =".Flycast.com/server/"

The numeral 89 indicates that this key word is the 89th in the list of key words for the required anchor targeter. Each targeter may be applied to the current block of HTML text sequentially and/or as an embedded sub-routine. If an unfinished HTML block of data is detected, the data receiving module 407 is notified by filter module 408 to re-send the unfinished block along with any new data available. Throughout the process if a target (HTML tag or keyword structure to be identified) is detected then the relevant targeter is told to filter its located target text according to its associated stored parameters as pre-configured prior to use of filter module 408 and data receiving module 407. A process of the type described in the example above is repeated until the data receiving module 407 informs filter module 408 that there is no more data for it to process.

Fig. 7 schematically illustrates a key word list of the type associated with a targeter of the type identified in Fig. 6. The list illustrated comprises single human English language words of an adult nature which have been pre-selected for detection by one or more targeters such that if a match is found then the targeter, during execution, is configured to modify the word such as by replacing it with white space or removing the entire contents of the HTML tag altogether for example. List 701 comprises various single words 702, 703 and comprises a total number M words of an adult nature as indicated by the last entry at 704. Fig. 8 schematically illustrates a similar list to that detailed in Fig. 7, the list 801 comprising associations of words configured to trigger execution of a given targeter to process any matched phrase found. Thus for example, the phrase at 802 "live sex" may be pre-set in the filter module 408 as a phrase to be deleted or overwritten with white space by a given targeter currently executing. Similarly, the phrase "Soho show" at 803 may be entered in the list to also effect the operation of one or more given targeters accordingly. Lists 701 and 801 may be held in a suitably configured data structure stored in memory 403 and accessible by filter module 408. Furthermore, lists 701 and 801 may be pre-set by a given manufacturer and/or modified according to a given user's particular requirements via a graphical user interface provided to enable a given user to modify the operation of filter module 408 as desired.

Further lists may be configured for identifying particular web sites referred to from a host site as being unsuitable or not required for a given user's requirements – such lists may be configured in a similar manner to lists 701 and 801 and are suitably configured as lists of web site addresses. These lists together with those shown in Figs. 7 and 8 can also be accessed semi independently with only a requirement for identification of the script begin and end command tags. Accordingly, within this main block which probably includes further command tags, the filter module can simply identify key words that signify the nature of the web site and hence apply filtering to the whole web page independent of the subsequent detection or lack thereof of any further tags within

the web page. This enables simple web pages using little HTML to be filtered in conjunction with pages using more complex combinations of HTML and advantageously allows combined filtering of advertisements together with site related content such as pornography or violence.

Fig. 9 further details the main steps executed by the data receiving module 407 identified in Fig. 4. The data receiving module 407 may be configured as a software entity and written in a suitable language such as Visual C ++ and/or Java. Module 407 is configured to wait for an incoming HTML page requested by a given user using a suitably configured browser presented to the user on screen 103. At step 901 the data receiving module is configured to wait for a HTML page whereafter at step 902 the module is triggered to receive a detected incoming HTML page. Following step 902, at step 903 the data receiving module is configured to read data received in the incoming HTML page and identify the host, URL and referrer of the newly received page. Following step 903, at step 904 the identified host, URL and referrer data are transmitted to filter module 408 for initialization purposes to configure all targeters of filter module 408 as required.

The filter module 408 may suitably be configured in a high level programming language such as Visual C++. It will be appreciated by those skilled in the art that the filter module 408 may be configured to operate and filter data using a variety of methods such as identifying tags and processing accordingly and for example identifying and relating particular tag types with particular key words stored in a list. Using the latter method if a tag/key word match is found the tag within the HTML information block being processed is filtered – this method therefore provides some flexibility in respect of filtering only tagged data comprising particular key words of a type deemed to be non-required by a given user. For advertisement type filtering this method is found to be particularly suitable.

Fig. 10 schematically illustrates one possible sequence of steps which may be undertaken by a suitably configured filter module 408 as configured in accordance with the present invention. At step 1001 filter module 408 is configured to receive the (next) HTML page of data received and transmitted via data receiving module 407. Following step 1001 the filter module 408 is configured to initialize each of the filter targeters such as those schematically illustrated in Fig. 6. Following step 1002 module 408 asks a question as to whether or not the current block of HTML data has been fully received. If the answer to the question at step 1003 is in the negative then control is passed to step 1004 and filter module 408 is configured to notify data receiving module 407 that the current block of HTML data must be re-sent. Following step 1004 control is therefore returned to step 1001 with steps 1001 – 1003 being repeated until the current block of data being processed is determined to be fully received at step 1003.

Following receipt of the complete HTML page, the question at step 1003 is therefore answered in the affirmative and control is passed to step 1005 and the first (or next) targeter is activated for operation. Following step 1005 control is passed to step 1006 wherein a question is asked as to whether any targets to be identified by the current targeter are present in the current block of HTML data. If the question asked at step 1006 is answered in the negative then control is returned to step 1005 and the next targeter is activated for operation. However if the question asked at step 1006 is answered in the affirmative then any located targets are processed by the current targeter at step 1007. Processing at step 1007 is further detailed in Fig. 11 and described below. Processing at step 1007 may include return of control to step 1005 under certain circumstances, as indicated by flow control line 1008, with steps 1005 – 1007 repeated.

Following completion of processing at step 1007, control is passed to step 1009 where a further question is asked as to whether any more targeters are to be applied to the current HTML page of data to be processed. If the answer to the question at step 1009 is answered in the affirmative then control is returned to

step 1005 and the next targeter is configured for execution. However, if the question asked at step 1009 is answered in the negative then control is passed to step 1010 where a further question is asked as to whether any more HTML pages have been received and buffered for processing. If the answer to the question at step 1010 is answered in the negative then processing is terminated at step 1011. However, if the question asked at step 1010 is answered in the affirmative then control is returned to step 1001 and the next HTML page received with processing steps 1001 – 1011 repeated accordingly.

Exemplary filter module steps 1001 – 1011 may comprise additional steps such as for example to process nested tag structuring which may be present in a given HTML page. Thus a given targeter may be configured to identify outer targets and/or inner targets for example. The steps may also include further steps to desirably deal with a variety of other potential situations as will be understood by those skilled in the art. Furthermore it is to be appreciated that it is possible to activate several mutually exclusive targeters in parallel to speed through the filtering process.

Fig. 11 further details a preferred exemplary sequence of steps involved in processing step 1007 of Fig. 10, this step being configurable in a variety of ways depending upon the types and level of filtering required. However, the example shown in Fig. 11 is included to provide a typical best mode example to the skilled person in the art for configuring operation of a given targeter. Following step 1006 control is passed to step 701 wherein a question is asked by the targeter as to whether or not the tagged data under consideration is to be modified, replaced or remain unchanged. If the answer to the question at step 1101 is answered in the negative then in this particular example the targeter is configured to delete the relevant tag type command information (such as a pop up window type tag or hyperlink for example) and control is returned to step 1005 wherein the next targeter is activated. However, if the question asked at step 1101 is answered in the affirmative then control is passed to step 1102 wherein the current targeter processing routine is applied to the targets (tagged structures) to which it is to

process. Following step 1102 control is passed to step 1103 wherein a further question is asked as to whether or not the targeter comprises instructions configured to check for any matchable data content held within the tagged structure. If the question asked as step 1103 is answered in the negative then control is returned to step 1005. However, if the question asked at step 1103 is answered in the affirmative then control is passed to step 1104 wherein filter module 408 is configured to read the data content associated with the tagged command, and in accordance with pre-configured rule based instructions, to compare the tagged content against stored listed data for any relevant data matches. Such lists may comprise stored lists in memory of the type described in Fig. 7 and 8 and/or lists of certain web-site addresses for example. Following step 1104 control is passed to step 1105 wherein a further question is asked as to whether or not a match has been found for the current type of tag structure under consideration. If the question asked at step 1105 is answered in the negative then control is returned to step 1104 and steps 1104 – 1105 are repeated until the question asked at step 1105 is answered in the affirmative to the effect that a "key word" match has been found. Following identification of a match the question asked at step 1105 is answered in the affirmative and control is passed to step 1106 wherein the particular matched data found is processed and, for example overwritten by white space. Following step 1106 control is passed to step 1107 wherein a further question is asked as to whether or not there is any further tagged content to be processed by the current targeter. If this question is answered in the affirmative then control is returned to step 1104 and steps 1104 to 1107 repeated. However, if the question asked at step 1107 is answered in the negative then control is returned to step 1010.

The present invention as described is thus able to filter both advertising based material and material of an adult nature and thus it will be appreciated that the invention comprises various components for filtering adult sites, some theme sites such as violence and bad language, and adverts which may otherwise appear on the user's screen 103. The invention may suitably be configured to automatically activate when a user opens a given web browser. The invention is

considered to provide various advantages over existing filters which are not integrated adult filters and advert filters and which do not attempt to read the information formed as part of the HTML document body. Thus, as well as removing a large proportion of web advertising download speeds are advantageously increased increasing productivity, and therefore on-line costs for a given web user are reduced. Not only is advertising material able to be filtered in accordance with a given user's requirements, but this filtering may be enhanced by examining strings of characters so as to filter out particular words and/or phrases may be present in certain adverts or other information components of a given HTML page.

The invention may also comprise an interactive on-line submissions facility which provides users with an immediate method for manually or automatically reporting any missed advertisements. Thus, when a new type of advertisement is reported, not only will the updated filter work for the reported site, but also on any site on the Internet using the same advertising method. Each time an advertisement slips through the existing filters and is reported by a user a new filtering mechanism may be created that will remove this and all similar advertisements. The invention may also incorporate an exceptions facility that allows filtering to be turned off, either for whole sites or for individual pages within a site. Both pop up window advertisements and banner advertisements may be filtered, pop up window advertisements being those that require the user to close the relevant windows before continuing. The invention advantageously may, by deletion of certain material, provide extra space for containing actual required data content and therefore the overall information density of a given received and required HTML page can be advantageously increased. While the invention is aimed at individual users, it may also be configured in a number of different versions of the basic filtering product and thereby aimed at different markets and adapted to reflect the needs of each particular user group. For example, private individuals may download a given version of the invention from the Internet or obtain the system by direct mail.

The invention may also comprise a privacy filter designed to prevent a given user's surfing habits of the web from becoming known to third parties. In certain embodiments of the invention the filter may require a download of less than 350K and it may be updated by means of a download of less than 20K (uncompressed). The invention may be configured with very few system dynamic link libraries (DLL's) and no plug-ins and may be configured to add zero system files when installed on a given user's computer.

Methods within the ambit of the present invention include detection of given mark up language tag type commands and subsequent processing and also processing of tag type commands under control of key word or of other text based matter detection. Filter module 408 may be configured to operate as a main filter with sub-routine calls being made to activate a plurality of sub-filters configured to filter particular types of text or image based material.

It is to be appreciated that pluralising behaviour can readily be added to matching procedures without difficulty. The pluralising function simply adds "s" to the original term and tests again and then adds "es" to the original term and tests again. For example, when searching for the term "leg", the pluralisation function creates terms such as "legs" and "leges" as possible further terms to be searched. Typically a f_Text filter containing a list of keywords will be searched.

The examples described are considered to be representative of best modes of carrying out the invention, but as indicated above it is possible to configure the filter module in a variety of ways to suit particular requirements. Thus, while the invention is described in some detail with specific reference to a single preferred embodiment and various alternatives there is no intent to limit the invention to the particular embodiment described or those specific alternatives. Thus, the true scope of the present invention is not to be considered as limited to any one of the foregoing described embodiments, but is instead defined by the appended claims.

Claims:

1. A client side data processing apparatus configurable for use with a computer system having a browser, said apparatus configured to process a block of information requested and received over a telecommunications network, said information comprising potentially required data content and tag type mark up language commands for controlling display of said potentially required data content by the browser, said apparatus comprising:

means for identifying a plurality of types of said received tag type commands; and

means configurable to process said received and identified tag type commands according to a pre-defined set of rules.

2. A data processing apparatus as claimed in Claim 1, wherein said identification means comprises means for reading said received information character by character and means for comparing a pattern of said characters with a pre-stored list of tag type command syntax.

3. A data processing apparatus as claimed in Claim 1 or 2, wherein said identification means comprises means for identifying a plurality of types of said tag type commands which are used for controlling display of electronic advertising information.

4. A data processing apparatus as claimed in any preceding claim, wherein said identification means comprises means for identifying tag type commands which specify a specific size of an electronic data banner to be displayed.

5. A data processing apparatus as claimed in any preceding claim, wherein said set of rules are adaptable according to a given user's requirements of said apparatus.

6. A data processing apparatus as claimed in any preceding claim, wherein said rules specify said processing to include modifying an identified tag type command according to a pre-defined criteria thereby changing its effect on execution.

7. A data processing apparatus as claimed in any preceding claim, wherein said rules specify said processing to include removing said tag type command from said received information.

8. A data processing apparatus as claimed in any preceding claim, wherein said rules specify said processing to include allowing the identified tag type command to be executed without any modification.

9. A data processing apparatus as claimed in any preceding claim, wherein said rules specify said processing to include replacing said identified tag type command with a stored tag type command.

10. A data processing apparatus as claimed in any preceding claim, wherein said configurable processing means may be modified according to a given user's preferences.

11. A data processing apparatus as claimed in any preceding claim, wherein said configurable processing means comprises further processing means arranged to identify and selectively process potentially non-required pre-defined data types in received data content associated with the identified tag type.

12. A data processing apparatus as claimed in Claim 11, wherein the further processing means is arranged to identify and selectively process potentially non-required pre-defined data types in the received data content in dependence of a URL of a web page relating to the currently received data.

13. A data processing apparatus as claimed in Claim 11 or 12, wherein said further processing means comprises:

means for reading the data comprising the content;

means for comparing the read content data with a stored list of potentially non-required data types to search for in the content data and identifying any matches found; and

means for processing the identified matched data in accordance with previously stored processing instructions associated with each potentially non-required data type in the list.

14. A data processing apparatus as claimed in Claim 13, wherein said list of stored potentially non-required data types comprises a list of human language words.

15. A data processing apparatus as claimed in Claim 13, wherein said list of stored potentially non-required data types comprises a list of certain groupings of human language words.

16. A data processing apparatus as claimed in any of Claims 13 to 15, wherein said means for processing said identified matched data includes means to prevent the display of said identified matched data.

17. A data processing apparatus as claimed in any of Claims 13 to 16, wherein said means for processing said identified matched data includes means to prevent the display of all of the data content associated with the identified tag type being considered, and including said identified matched data.

18. A data processing apparatus as claimed in any preceding claim, wherein said received information comprises a Hypertext Mark-Up Language (HTML) page.

19. In a computer system having a browser for displaying requested information received over a telecommunications network, said information comprising potentially required data content and tag type mark up language commands for controlling said display function, a method of controlling the functionality of said received tag type commands, said method comprising:

identifying a received tag type command;

processing said identified tag type command according to a pre-defined set of rules configured for application to tag type commands of said identified type.

20. The method as claimed in Claim 19, wherein said step of identification comprises comparing said received tag type command with a pre-stored list of tag type commands and identifying a match.

21. The method as claimed in Claim 20, wherein if a match cannot be found, the method further comprises saving details of the tagged command under consideration and providing a warning message to the user of said system.

22. The method as claimed in Claim 19 or 20, wherein following a given tagged command having been identified, said processing step includes loading executable processing instructions associated with said tag type and processing said tag type accordingly.

23. The method as claimed in any of Claims 19 to 22, wherein said processing step includes one of:

ignoring said tag type command; enabling said tag type command to execute in its original form;

replacing said tag type command with a pre-set replacement command; or

modifying said tag type command according to pre-defined stored rules for said tag type command thereby changing the executable effect of said identified tag type command.

24. The method as claimed in any of Claims 19 to 23, wherein said processing step includes:

reading the data comprising the data content associated with the tag;

comparing the read content data with a list of previously stored potentially non-required data types to search for in the content data and identifying any matches found;

configuring processing means with previously stored processing instructions associated with each potentially non-required data type in the list; and

processing one or more the identified matched data in accordance with the associated instructions.

25. The method as claimed in Claim 24, wherein said list of stored potentially non-required data types comprises a list of human language words.

26. The method as claimed in Claim 24, wherein said list of stored potentially non-required data types comprises a list of certain pre-defined groupings of human language words.

27. The method as claimed in any of Claims 24 to 26, wherein said step of processing in accordance with said associated instructions comprises preventing display of said identified matched data.

28. The method as claimed in any of Claims 24 to 27, wherein said step of processing in accordance with said associated instructions includes preventing the display of all of the data content associated with the identified tag type being considered, and including said identified matched data.

29. The method as claimed in any of Claims 19 to 28, wherein said received information comprises a HTML page.

30. A method of filtering non-required data from information received over a telecommunications channel, the method comprising identifying a received tag type command from within the received information, scanning the data content specifically associated with the received tag type command and filtering at least a portion of the data content in response to matching an item in a pre-stored data list with the portion of the data content.

31. A method according to Claim 30, wherein the data content which is the subject of the scanning and filtering steps comprises textual data.

32. A method as claimed in Claim 30 or 31, wherein there are a plurality of pre-stored data lists each being specifically associated with at least one tag type command and the matching step comprises searching those lists associated with the received tag type command.

33. A method as claimed in any of Claims 30 to 32, wherein the filtering step comprises selectively filtering the portion of the data content without filtering the entire data content associated with the received tag type command.

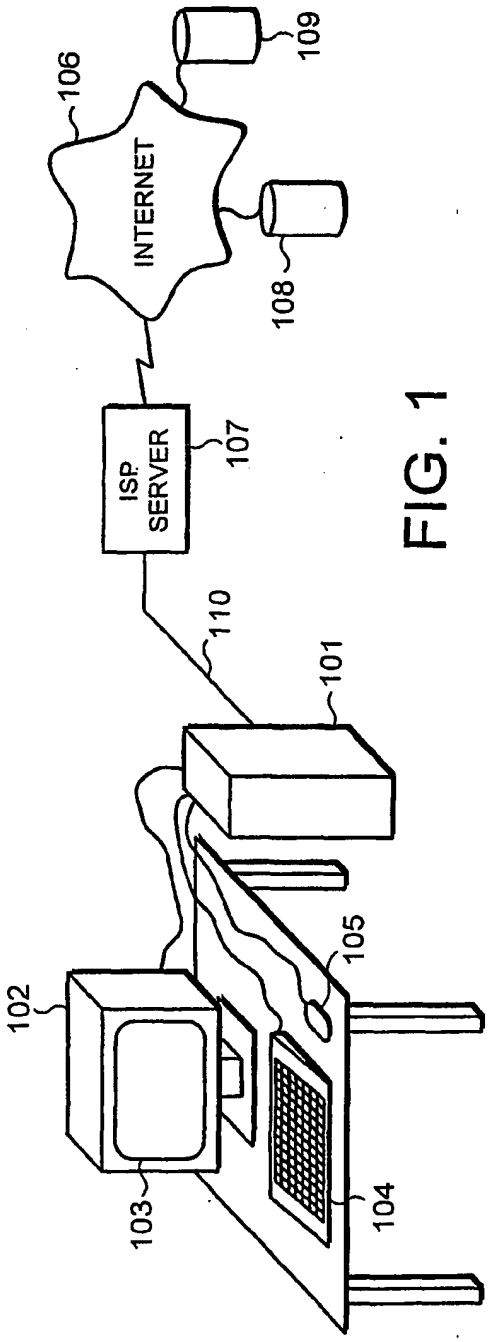


FIG. 1

```
<HTML>~202

<HEAD>~204   209   206   208        205
207
<TITLE> No1 Webfilter, INC. </TITLE> </HEAD>

<BODY>~210
      213                                          214
<HI> Welcome to No1 Webfilter's Home Page </HI

<IMG SRC = "http://www.webfilter.com/images/logo.gif"~217
     ALT = "No1 Webfilter logo"> <BR>
   218                          219
We are <B> very pleased </B that you have taken the time
to visit us....
      225
<P> A link to further information about our new filter product
is provided below. <HR>
      215                          216
<H2> New filter-information </H2>
      223    221
212 <UL> <AHREF = "http://webfilter.com/products/new">
                                   220
New filter </A>
             222
</UL>~224

<H2> Telephone No </H2>

<UL> <L1> By telephone: 0800 11 22 33
     <L1> By fax:       0800 11 22 34

</UL>

</BODY>~211

</HTML>~203                                    201
```

FIG. 2

Welcome to No1 Webfilter's Home Page
303
We are very pleased that you have taken the time to visit us.
A link to further information about our new filter product is provided below.
New filter-information
• New Filter ~302
Telephone No.
• By telephone: 0800 11 22 33
By fax: : 0800 11 22 34
103
301

FIG. 3

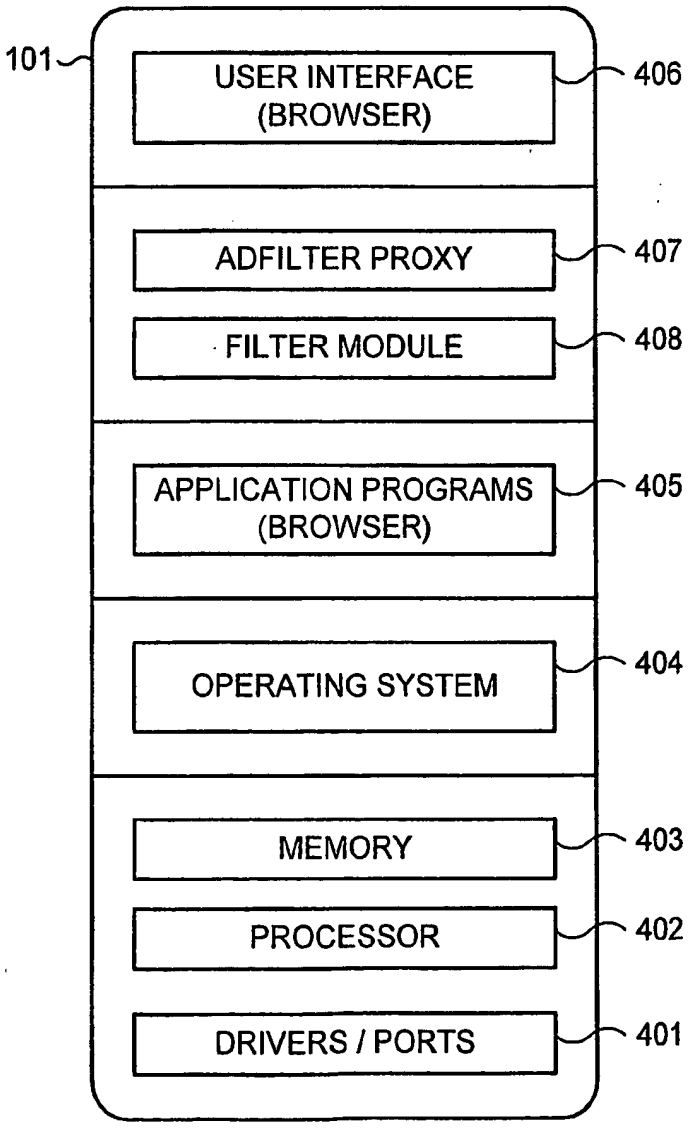

101
USER INTERFACE (BROWSER)
406
ADFILTER PROXY
407
FILTER MODULE
408
APPLICATION PROGRAMS (BROWSER)
405
OPERATING SYSTEM
404
MEMORY
403
PROCESSOR
402
DRIVERS / PORTS
401

FIG. 4

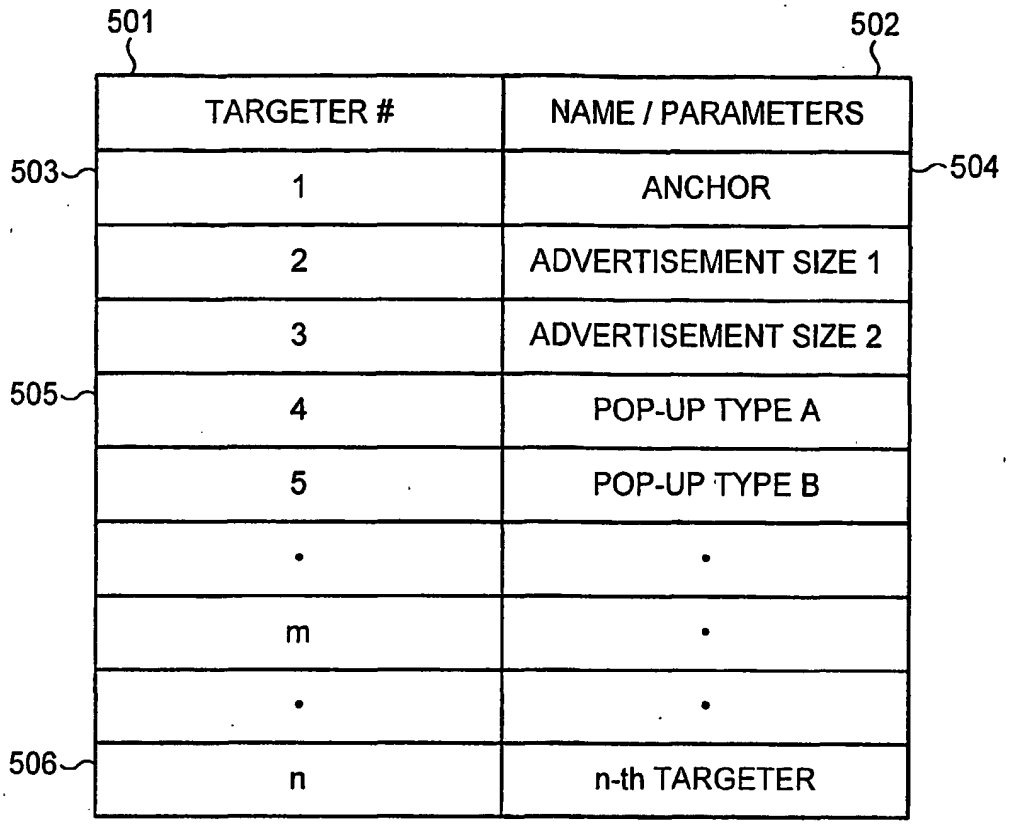

501
502
TARGETER #
NAME / PARAMETERS
503
504
1
ANCHOR
2
ADVERTISEMENT SIZE 1
3
ADVERTISEMENT SIZE 2
505
4
POP-UP TYPE A
5
POP-UP TYPE B
m
506
n
n-th TARGETER

FIG. 5

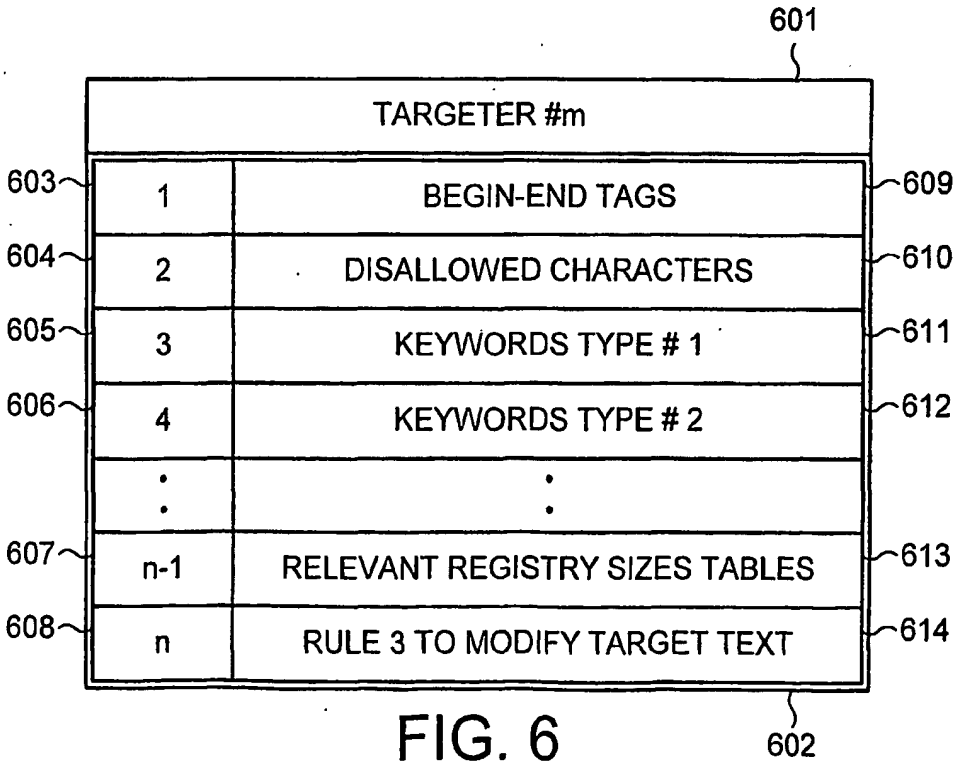

601
TARGETER #m
603
609
1
BEGIN-END TAGS
604
610
2
DISALLOWED CHARACTERS
605
611
3
KEYWORDS TYPE # 1
606
612
4
KEYWORDS TYPE # 2
607
613
n-1
RELEVANT REGISTRY SIZES TABLES
608
614
n
RULE 3 TO MODIFY TARGET TEXT
602

FIG. 6

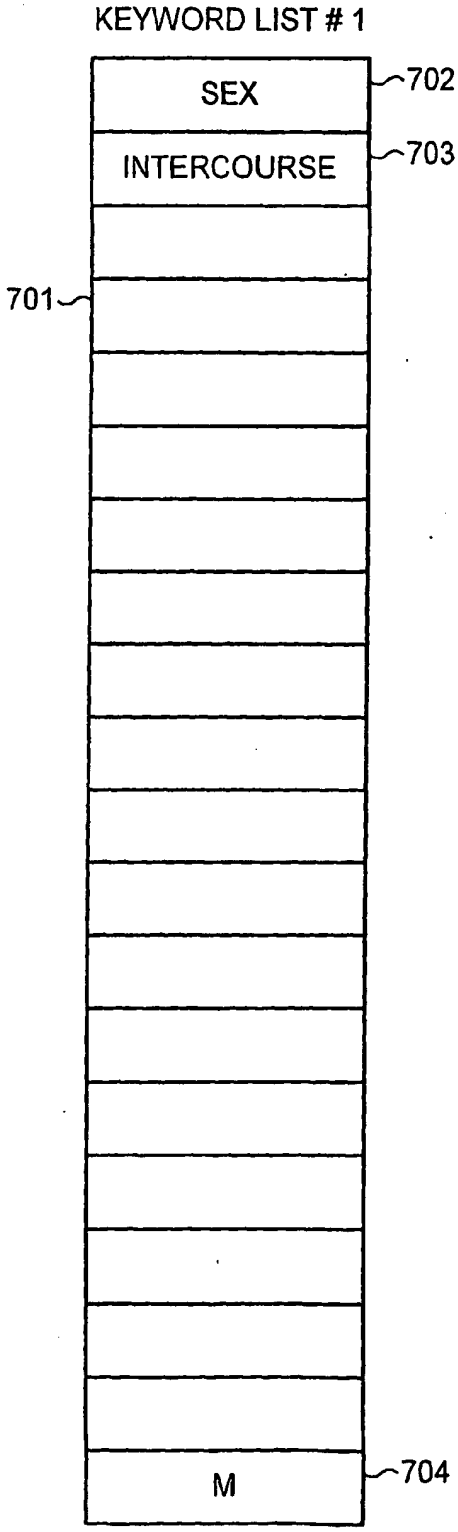

KEYWORD LIST # 1
SEX
INTERCOURSE
M
701
702
703
704

FIG. 7

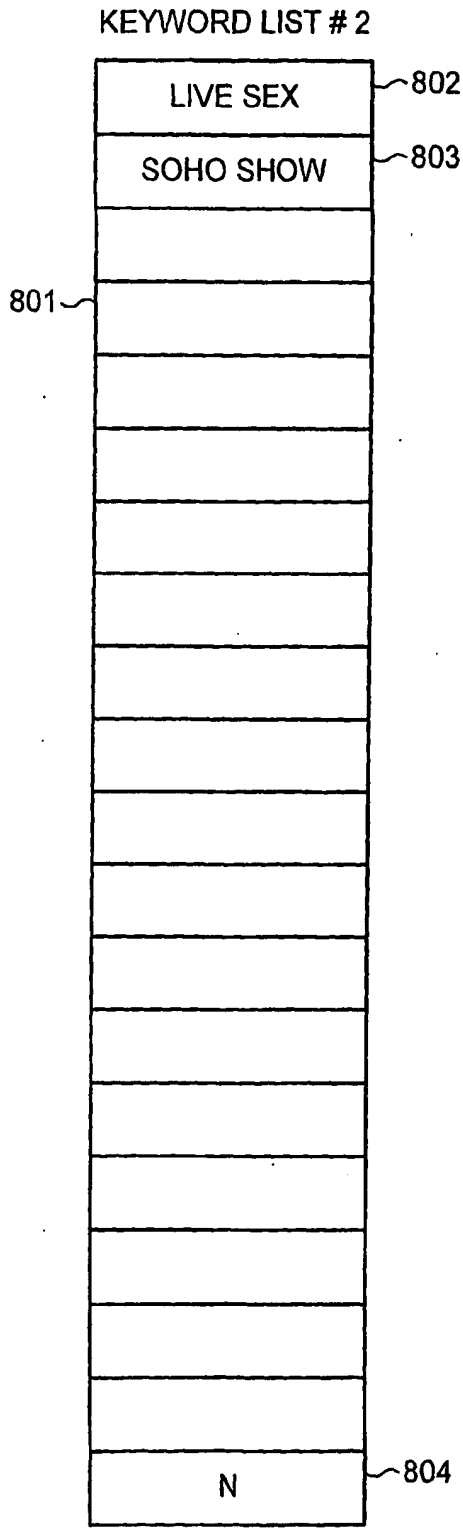

KEYWORD LIST # 2
LIVE SEX
SOHO SHOW
N
801
802
803
804

FIG. 8

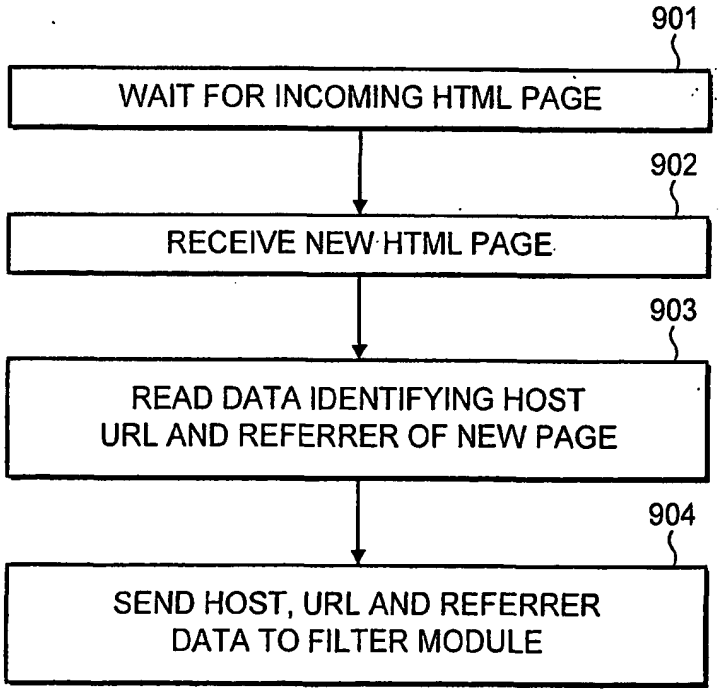

901
WAIT FOR INCOMING HTML PAGE
902
RECEIVE NEW HTML PAGE
903
READ DATA IDENTIFYING HOST URL AND REFERRER OF NEW PAGE
904
SEND HOST, URL AND REFERRER DATA TO FILTER MODULE

FIG. 9

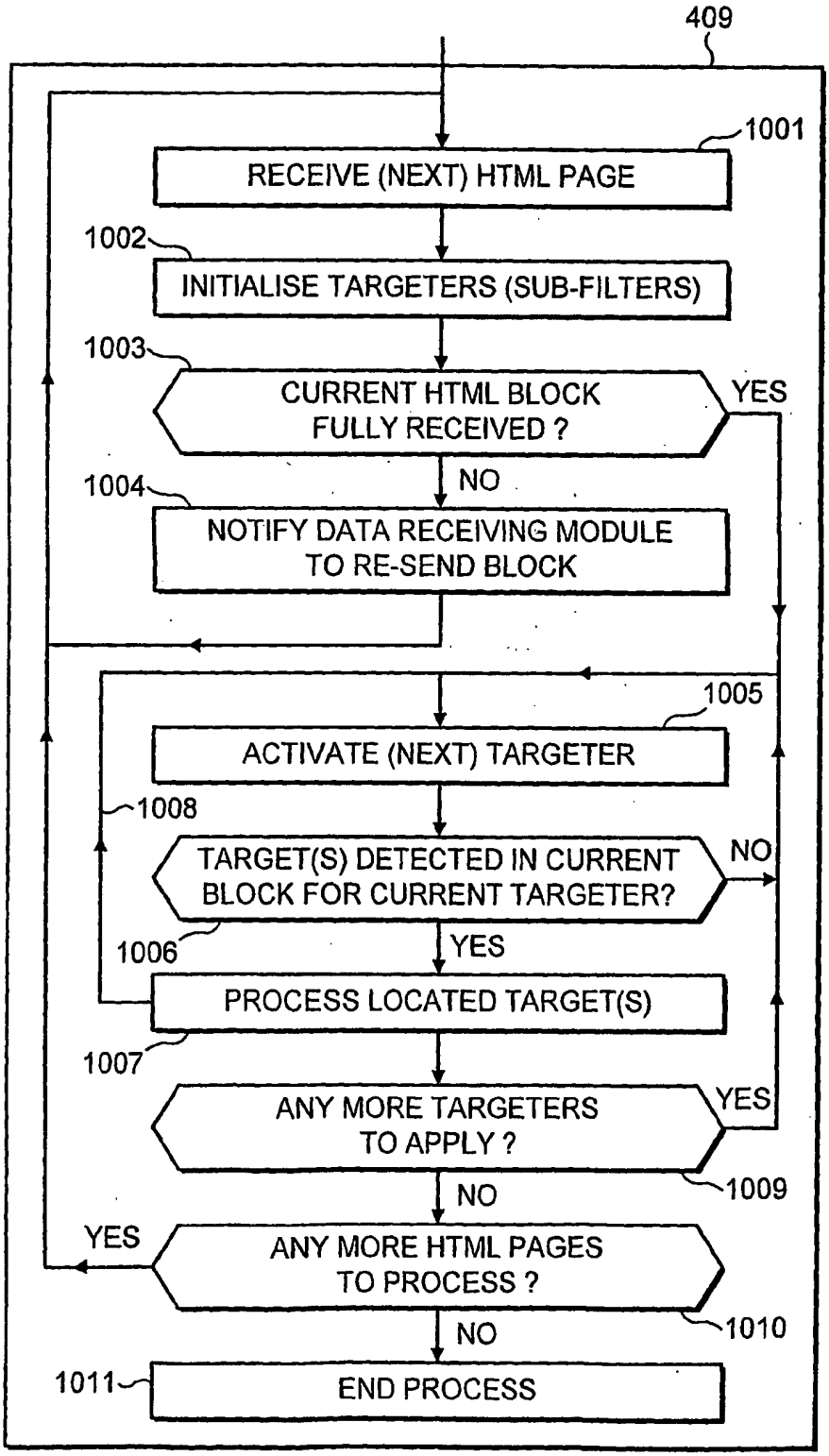

409
1001
RECEIVE (NEXT) HTML PAGE
1002
INITIALISE TARGETERS (SUB-FILTERS)
1003
CURRENT HTML BLOCK FULLY RECEIVED ?
YES
NO
1004
NOTIFY DATA RECEIVING MODULE TO RE-SEND BLOCK
1005
ACTIVATE (NEXT) TARGETER
1008
TARGET(S) DETECTED IN CURRENT BLOCK FOR CURRENT TARGETER?
NO
1006
YES
PROCESS LOCATED TARGET(S)
1007
ANY MORE TARGETERS TO APPLY ?
YES
1009
NO
YES
ANY MORE HTML PAGES TO PROCESS ?
1010
NO
1011
END PROCESS

FIG. 10

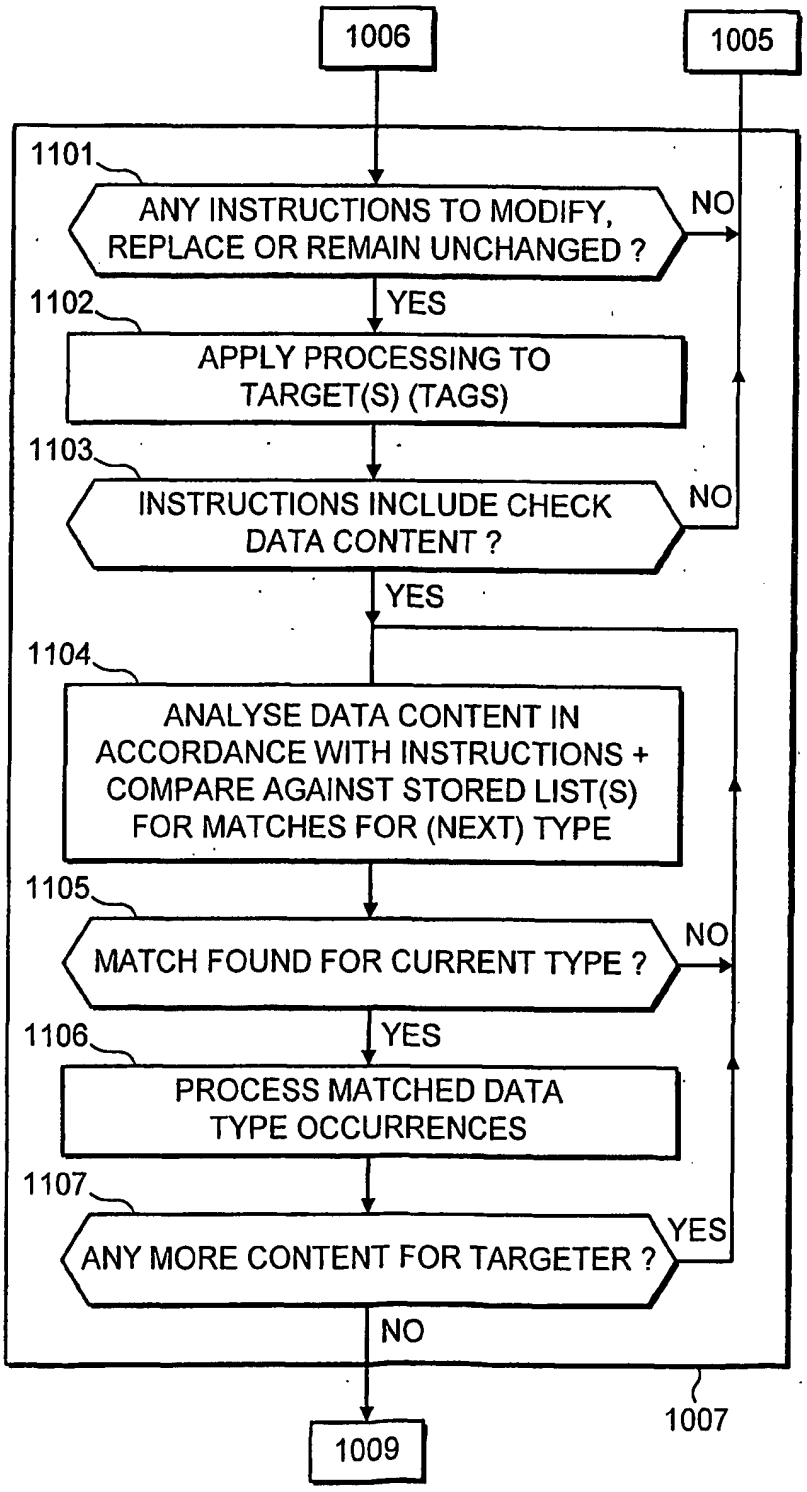

1006
1005
1101
ANY INSTRUCTIONS TO MODIFY, REPLACE OR REMAIN UNCHANGED ?
NO
YES
1102
APPLY PROCESSING TO TARGET(S) (TAGS)
1103
INSTRUCTIONS INCLUDE CHECK DATA CONTENT ?
NO
YES
1104
ANALYSE DATA CONTENT IN ACCORDANCE WITH INSTRUCTIONS + COMPARE AGAINST STORED LIST(S) FOR MATCHES FOR (NEXT) TYPE
1105
MATCH FOUND FOR CURRENT TYPE ?
NO
YES
1106
PROCESS MATCHED DATA TYPE OCCURRENCES
1107
ANY MORE CONTENT FOR TARGETER ?
YES
NO
1007
1009

FIG. 11